

附件 4:

中汽协会《自动驾驶感知训练数据集标注格式要求》

团体标准编制说明

一、工作简要过程

(一) 任务来源

1、项目立项背景

本标准立项编制的背景主要基于以下几个方面的考虑：

- (1) 技术发展的需要：随着自动驾驶技术的快速发展，感知模型作为自动驾驶系统的核心组成部分，其性能和准确性直接影响到自动驾驶车辆的安全性和稳定性。然而，不同的感知模型可能采用不同的输出格式，这导致在自动驾驶系统的集成和测试过程中存在较大的不便。因此，制定统一的感知模型输出格式标准，有助于推动自动驾驶技术的标准化和规范化发展；
- (2) 行业协作的需求：自动驾驶技术的研发涉及多个领域，包括车辆工程、计算机视觉、人工智能等。不同的研发团队和机构可能采用不同的感知模型和数据处理方式。制定统一的感知模型输出格式标准，可以促进不同团队之间的协作和交流，加速自动驾驶技术的研发进程；
- (3) 安全性的保障：自动驾驶车辆的安全性是公众最为关心的问题之一。感知模型作为自动驾驶系统的重要组成部分，其输出格式的准确性和一致性直接影响到车辆对环境的感知和决策。制定统一的感知模型输出格式标准，有助于确保感知数据的准确性和可靠性，从而提高自动驾驶车辆的安全性；
- (4) 法规和标准的要求：随着自动驾驶技术的不断发展，各国政府和相关机构也在逐步制定和完善相关的法规和标准。制定统一的感知模型输出格式标准，有助于满足法规和标准的要求，推动自动驾驶技术的合法化和规范化应用。

2、计划任务编号

(二) 主要起草单位及任务分工

上海机动车检测认证技术研究中心有限公司：标准编制工作组牵头单位，全面负责标准项目的立项、标准文本以及编制说明的编写、标准编制工作的组织及协调等工作。

其他对于标准编制有重要贡献的单位包括（不分先后）：智行众维、上海国际汽车城集团、如祺出行、小马智行、毫末智行、长安安驿、AutoCore、文远知行、广汽研究院、蔚来汽车。

(三) 标准研讨情况

在本标准立项、起草的过程中，召开了多次调研、讨论等会议，以收集行业意见、梳理技术难点、明确标准框架和内容。以下是会议情况的介绍及阶段性成果的突出展示：

(1) 调研会议情况

行业调研会议：

邀请了来自汽车制造、自动驾驶研发、计算机视觉等领域的专家和企业代表，就当前自动驾驶感知模型的应用现状、输出格式的差异及挑战进行了深入交流。通过调研，我们了解

到不同企业和团队在感知模型输出格式上存在的差异和痛点,为后续标准制定提供了重要的参考依据。

技术调研会议:

针对感知模型的技术原理、算法特点、数据处理流程等进行了技术调研。会议邀请了相关领域的专家学者,就感知模型输出格式的标准化需求、可行性及潜在问题进行了深入探讨。通过技术调研,我们明确了标准制定的技术基础和方向。

二、讨论会议情况

标准框架讨论会议:

在标准框架讨论会议上,与会人员就标准的结构、章节划分、内容要点等进行了充分讨论。经过多轮讨论和修改,最终确定了标准的初步框架,为后续内容的编写奠定了基础。

内容细节讨论会议:

在内容细节讨论会议上,重点对感知模型输出格式的具体要求、参数定义、数据格式等进行了深入讨论。与会人员积极发言,提出了许多宝贵的意见和建议。经过反复讨论和修改,逐步完善了标准的内容细节。

三、阶段性成果

明确了标准制定的目标和原则:

通过调研和讨论,我们明确了自动驾驶感知模型输出格式要求标准制定的目标和原则,即推动自动驾驶技术的标准化和规范化发展,提高感知数据的准确性和可靠性,保障自动驾驶车辆的安全性。

形成了初步的标准框架和内容:

经过多轮讨论和修改,我们形成了初步的标准框架和内容。标准包括引言、术语定义、感知模型输出格式要求、测试验证方法等多个章节,对感知模型输出格式的具体要求进行了详细规定。

收集了广泛的行业意见和建议:

通过调研和讨论会议,我们收集了来自不同领域和行业的意见和建议,为标准的制定提供了丰富的参考依据。同时,我们也与行业内的企业和团队建立了良好的合作关系,为后续标准的推广和应用奠定了基础。

二、标准编制原则和主要内容

1、标准编制原则

1) 规范性原则:本标准按照 GB/T 1.1-2020《标准化工作导则第一部分:标准的结构和编写》的规定和要求编制,在标准框架、结构和内容等方面符合要求。

2) 科学实用原则:标准内容全面,采纳多方面实际经验,适应汽车企业在自动驾驶感知模型训练的实际需要。

2、主要内容介绍

标准主要针对自动驾驶领域的感知模型训练数据集格式提出规范要求,用于规范数据集的文件组织结构、字段与标识名称、数据类型与格式,以及部分精度要求。具体包括以下几个模块内容:

1) 时空参考系基准:规定了数据集中每条记录生成的时间表示及时间同步机制,并定义物体在三维以及路面、图像内的二维空间的描述与表示方法。

2) 数据集文件结构:说明一个自动驾驶感知训练数据的文件需要包含的数据文件类型,包括采集条件、采集数据、数据标注、类别属性四类,并说明各个文件间的索引与依赖关系。

3) 数据文件格式与字段命名规范:明确数据文件所采用的数据记录格式,并说明各个字段的命名及取值表示规范。

4) 实例类别命名规范与标注重要性：对常见感知实例的命名进行规范，明确多种感知实例的类别从属关系，并规定感知实例的感知重要性。

三、采用国际标准和国外先进标准情况

本标准具有一定的前瞻性，在制定过程中未能找到可采用的标准。

四、主要关键指标及试验验证情况

本标准在制定过程中，已完成汽车行业自动驾驶技术常用数据集记录格式、数据记录类型的调研工作。对公开数据集、多家国内自动驾驶领域龙头企业的数据集构建方案进行整理与分析，根据数据集的知名度、通用性，对数据集格式进行整合与协调。主要参考方面如下：

1) 公开数据集：通过线上文档查阅的方式，完成对 Waymo、nuSense 等公开数据集记录格式的调研工作，完成不同类型数据记录的必要性与基本格式的初步构建。

2) 企业自建数据集：通过与企业相关负责人沟通会谈的方式，对标准中相关要求、描述进行符合性确认，明确标准内容与各企业实际应用的差异，并以差异最小化原则，对标准内容进行适配修改，使得标准更贴近企业实际生产需要。

五、与现行法律、法规和政策及相关标准的协调性

本标准与现行国家法律、法规，现行标准，制定中标准均无任何相悖之处。

六、贯彻标准的要求和措施建议

本标准发布后，建议由数据分会协调组织宣贯会议进行行业宣贯。

建议国内各汽车厂参照本标准要求，完成对感知模型的数据处理模块、现有感知训练数据集格式的调整。

七、其他需要说明的事项

无其他需要说明事项。